

AW
9-15-08

10/729,126

-34-

corresponding with a particular candidate English collocation. Generally, the Chinese collocations can be ranked from most to least probable.

At step 622, similarity information for
5 candidate English language collocations is calculated or generated using the feature vectors. In some embodiments, the cosine method is used to calculate similarities. However, other known methods of calculating similarity can be used as described
10 above.

At step 624, English language collocations having a similarity value exceeding a selected threshold are selected as synonymous collocations. In some embodiments, the selected threshold can differ
15 for collocations having different relation-types.

At step 626, a lexical knowledge base is augmented with the generated or selected synonymous collocations that can be used later in desired applications such as language generation.

FIG. 7 illustrates method 700 of
20 generating language using the lexical knowledge base constructed by another aspect of the present invention. At step 702, a lexical knowledge base having stored synonymous collocations is accessed,
25 obtained or received from any of the input devices described above or from any of the data storage devices described above. An input sentence is received at step 704. At step 706, the input sentence is parsed in order to recognize collocations as
30 described above.

AW
9-15-08